

A VOICE COMMAND IDENTIFIER FOR A VOICE RECOGNITION SYSTEMRelated Applications

[0001] This application is a continuation application, and claims the benefit under 35 U.S.C. §§120 and 365 of PCT application No. PCT/KR02/00268 filed on February 20, 2002 and published on September 26, 2002, in English, which is hereby incorporated by reference herein.

Background of the InventionField of the Invention

[0002] The present invention relates to a voice command identifier for a voice recognition system, especially to a voice command identifier for recognizing a valid voice command of a user by identifying user's voice command from a sound output from an embedded sound source.

Description of the Related Technology

[0003] It is generally known that a conventional voice recognition system can recognize a voice command spoken by a human effectively through a various kinds of methods (Detailed descriptions on the conventional recognizing methods or structures of the conventional voice recognition systems are already known in the art of the present invention, and are not direct subject matters of the present invention, so that they are omitted for simplicity.).

[0004] However, as shown in Fig. 1, a conventional home appliance 10, such as televisions, audio players or video players, which can produce a sound output, can not distinguish user's voice command from input sound, which was output by its own embedded sound source and re-input into itself by reflection and/or diffraction. Therefore, it is impossible to use the conventional voice recognition system for an apparatus with a sound source because the voice recognition system can not distinguish a voice command from a re-input sound.

[0005] A conventional approach for solving this problem eliminates a re-input sound from a received signal of a microphone 104 by estimating output sound with time. Let the received signal of the microphone 104 be $S_{mic}(t)$, and the sound signal output by a speaker 102 be $S_{org}(t)$. Then, the received signal of the microphone 104 $S_{mic}(t)$ includes a voice command signal $S_{command}(t)$ of a voice command spoken by a user and a distortion signal $S_{dis}(t)$ which is a distorted signal of the sound signal $S_{org}(t)$ by reflection and/or diffraction in its way to the microphone 104 from the speaker 102. This is expressed by Equation 1, as follows:

[Equation 1]

$$\begin{aligned} S_{mic}(t) &= S_{command}(t) + S_{dis}(t) \\ &= S_{command}(t) + \sum_{k=0}^N (A_k \times S_{org}(t - t_k)) \end{aligned}$$

[0006] Here, t_k is a delay time due to reflection and has a value of reflection distance divided by the velocity of sound. A_k (“environmental variable”) is a variable influenced by its environment and determined by the amount of energy loss of the output sound due to the reflection. Since output sound $S_{org}(t)$ is already known, it was asserted to be possible to extract user’s voice command only by determining values of A_k and t_k . However, it is very difficult to embody a hardware or a software system which can perform the direct calculations of the above Equation 1 in real time since the amount of calculation is too big.

[0007] There was another approach to decrease the amount of calculation by transforming the distortion signal $S_{dis}(t)$ with, for example, Fourier Transformation. But, it is required to know all environmental variables according to its real operating environment in advance, which is impossible.

Summary of Certain Inventive Aspects of the Invention

[0008] One aspect of the invention provides a voice command identifier which can perform the required calculation by decreasing the amount of calculations by acquiring and storing environmental variables on initial installation.

[0009] Another aspect of the invention provides a voice command identifier which is adaptive to change of environment by acquiring and renewing environmental variables when the system is placed under a new environment.

[0010] Another aspect of the invention provides a voice command identifier for a voice-producible system having an internal circuitry performing a predetermined function, an audio signal generator for generating a sound signal of audio frequency based on a signal provided from the internal circuitry, a speaker for outputting the sound signal as an audible sound, a microphone for receiving external sound and converting them into an electrical signal and a voice recognizer for recognizing an object signal included in the electrical signal from the microphone, including: a memory of a predetermined storing capacity; a microprocessor for managing the memory and generating at least one control signal; a first analog-to-digital converter for receiving the sound signal from the audio signal generator and converting them into a digital signal in response to control of the microprocessor; an adder for receiving the electrical signal from the microphone and outputting the object signal, which is to be recognized by the voice recognizer in response to control of the microprocessor; a second analog-to-digital converter for receiving the object signal and converting them into a digital signal; a first and second digital-to-analog converters for respectively converting retrieved data from the memory into analog signals in responsive to control of the microprocessor; and an output selecting switch for selecting one of outputs out of the second digital-to-analog converter and the audio signal generator in responsive to control of the microprocessor.

[0011] Another aspect of the invention provides a voice command identifying method for a voice-producible system having an internal circuitry performing a predetermined function, an audio signal generator for generating a sound signal of audio frequency based on a signal provided from the internal circuitry, a speaker for outputting the sound signal as an audible sound, a microphone for receiving external sound and converting them into an electrical signal and a voice recognizer for recognizing an object signal comprised in the electrical signal from the microphone, the method comprising: (1) determining whether a setting operation or a normal operation is to be performed; in case the determination result of the step (1) shows that the setting operation is to be performed, (1-1) outputting a pulse of a predetermined amplitude and width; and (1-2) acquiring an environmental coefficient uniquely determined by installed environment by digitizing a signal input into the microphone for a predetermined time period after the pulse is output; in case the determination result of the step (1) shows that the normal operation is to be

performed, (2-1) acquiring a digital signal by analog-to-digital converting a signal output from the audio signal generator; (2-2) multiplying the digital signal acquired by the step (2-1) with the environmental coefficient and accumulating a multiplied result; and (2-3) digital-to-analog converting an accumulated result into an analog signal and generating the object signal by subtracting the analog signal from the electrical signal output from the microphone.

Brief Description of the Drawings

[0012] Fig. 1 shows a schematic diagram of a space where a home appliance including a voice command identifier according to an embodiment of the present invention.

[0013] Fig. 2 shows a voice recognition system including a voice command identifier according to an embodiment of the present invention.

[0014] Fig. 3 shows a schematic diagram of a memory structure managed by the voice command identifier shown in Fig. 2.

[0015] Fig. 4 shows a flowchart of operation of the voice command identifier shown in Fig. 2 according to an embodiment of the present invention.

[0016] Fig. 5 shows a flowchart of a “setting operation” shown in Fig. 4 according to an embodiment of the present invention.

[0017] Fig. 6 shows a flowchart of a “normal operation” shown in Fig. 4 according to an embodiment of the present invention.

[0018] Fig. 7 shows waveforms of a test signal output during the normal operation shown in Fig. 6 and a received signal resulted from the test signal.

[0019] Fig. 8 shows waveforms of a sound signal output during the normal operation shown in Fig. 6 and a received signal resulted from the sound signal.

[0020] Fig. 9 shows a waveform of an output signal output during the normal operation shown in Fig. 6.

Detailed Description of Certain Embodiments of the Invention

[0021] Now, a voice command identifier according to embodiments of the present invention is described in detail with reference to the accompanying drawings.

[0022] Fig. 2 shows a voice recognition system including a voice command identifier according to an embodiment of the present invention. As shown in Fig. 2, the voice command identifier 100 may be provided to a voice-producible system (simply

called as a “system”, hereinafter), such as a television, a home or car audio player, a video player, etc., which can produce a sound output in itself. The voice-producible system having the voice command identifier 100 may include an internal circuitry 106 performing a predetermined function, an audio signal generator 108 for generating a sound signal $S_{org}(t)$ of audio frequency based on a signal provided from the internal circuitry 106, a speaker 102 for outputting the sound signal as an audible sound, a microphone 104 for receiving external sound and converting them into an electrical signal $S_{mic}(t)$, and a voice recognizer 110 for recognizing an object signal $S_{command}(t)$ included in the electrical signal $S_{mic}(t)$ from the microphone 104. The above described structure of the voice-producible system and its elements are known to an ordinary skilled person in the art of the present invention, so details of them are omitted for simplicity.

[0023] As described above about the conventional systems, the sound output by the system is re-input into the system by reflection or diffraction by various obstacles in the place where the system is located (see Fig. 1). Therefore, it is of very high probability that the voice recognizer 110 malfunctions because it can not distinguish a user’s command from the re-input sound of the same or similar pronunciation, wherein the re-input sound is output by the system itself and reflected or diffracted by the environment.

[0024] The voice command identifier 100 identifies the user’s voice command from the sound of the same or similar pronunciation included in the sound output by the system, and lets only the identified user’s voice command to be input into the voice recognizer 110 of the system.

[0025] The voice command recognizer 100 according to an embodiment of the present invention includes a first analog-to-digital converter 112 for receiving the sound signal $S_{org}(t)$ from the audio signal generator 108 and converting them into a digital signal, an adder 118 for receiving the electrical signal $S_{mic}(t)$ from the microphone 104 and outputting an object signal $S_{command}(t)$, which is to be recognized, and a second analog-to-digital converter 120 for receiving the object signal $S_{command}(t)$ and converting them into a digital signal.

[0026] The first and second analog-to-digital converters 112 and 120 perform their operations in response to control of a microprocessor 114 provided to the voice command identifier 100 of the present invention. The microprocessor 114 performs required calculations and control operations for controlling operations of the above described elements 112, 118 and 120, besides. The microprocessor 114 is one of the

general-purpose hardware and can be clearly defined by its operations described by this specification in detail. Other known details about microprocessors are omitted for simplicity.

[0027] The voice command identifier 100 may further include a memory (not shown) of a predetermined storing capacity. The memory may preferably be an internal memory of the microprocessor 114. Of course, an additional external memory (not shown) may be used for more sophisticated control and operation. Note that data converted into/from the sound signal is retrieved or stored from/into the memory according to control of the microprocessor 114. As for the type of the memory, it is preferable to use both volatile and nonvolatile types of memories, as described later.

[0028] The voice command identifier 100 further includes a first and second digital-to-analog converters 116 and 122 for converting retrieved data from the memory into an analog signal according to control of the microprocessor 114. The voice command identifier 100 further includes an output selecting switch 124 for selecting one of outputs out of the second digital-to-analog converter 122 and the audio signal generator 108 according to control of the microprocessor 114.

[0029] As shown in the drawing, the adder 118 performs subtraction operation of the output signal received from the first digital-to-analog converter 116 from the electrical signal $S_{mic}(t)$ from the microphone 104.

[0030] Fig. 3 shows a schematic diagram of a memory structure managed by the voice command identifier shown in Fig. 2. As shown in Fig. 3, the memory may be structured to have four (4) identifiable sub-memories 300, 302, 304 and 306. The first and second sub-memories 300 and 302 store data of an environmental coefficient $C(k)$, which is digitized one corresponding to the environmental variable A_k in the Equation 1. The environmental coefficient $C(k)$ reflects physical amount of attenuation and/or delay due to the environment in which the sound output by the speaker 102 is reflected and/or diffracted and re-input into the microphone 104. Therefore, as described later, even in case the sound signal $S_{org}(t)$ output by the system is changed by the characteristic nature of the environment where the system is installed, the user's voice command, which should be the object of recognition, can be distinguished from re-input sound, which is output by the system itself, by acquiring the environmental coefficient $C(k)$ through a setting procedure performed at the time of the first installation of the system at a specific environment.

[0031] It is preferable to use a nonvolatile memory as the first sub-memory 300 and a fast volatile memory as the second sub-memory 302. Therefore, the second sub-memory 302 may not be used in case processing speed is not important, or the first sub-memory 300 may not be used in case power consumption is not important.

[0032] The third sub-memory 304 sequentially stores digital signal $M(k)$'s, which is sequentially converted from the sound signal $S_{org}(t)$ from the audio signal generator 108. The third sub-memory 304, as described later, does not replace a value acquired by the prior processing operation with new value acquired by the present processing operation at the same storage area. The third sub-memory 304 stores every and each value acquired by several processing operations during a predetermined period on a series of storage areas until a predetermined number of values are acquired, where the storage area is shifted by one value and another. (This storage operation of a memory is called as "Que operation", hereinafter.) The Que operation of the third sub-memory 304 may be performed according to control of the microprocessor 114, or by a memory device (not shown) structured to perform the Que operation.

[0033] The fourth sub-memory 306 sequentially stores digital signals $D(k)$ into which the signal $S_{command}(t)$ ("object signal") output by the adder 118 is converted by the second analog-to-digital converter 120. It is also preferable to use a fast volatile memory as the fourth sub-memory 306. The third sub-memory 304 is used for the normal operation, and the fourth sub-memory 306 is used for the setting operation, as described later. Thus, it is possible to embody the third and fourth sub-memories 304 and 306 by only one physical memory device.

[0034] It is enough to distinguish the first to fourth sub-memories 300, 302, 304 and 306 from one another logically, thus it is not always necessary to distinguish them from one another physically. Therefore, it is possible to embody the sub-memories with one physical memory device. This kind of structuring memory device is already known to an ordinary skilled person in the art of the present invention, and detailed description on that is omitted for simplicity.

[0035] Now, referring to Figs. 4 to 9, operation of the voice command identifier 100 is described in detail. Fig. 4 shows a flowchart of operation of the voice command identifier shown in Fig. 2 according to an embodiment of the present invention. When power is applied to the system and the operation is started, the voice command identifier 100 determines to perform a setting operation (step S402). It is preferable to perform the step S402 when the setting operation has never been performed or when the

user wants to do it. Therefore, it is preferable to set the voice command identifier 100 to automatically perform a normal operation (refer to step S406), and to perform the setting operation (step S402) only when, for example, the user presses a predetermined button or a predetermined combination of buttons of the system. In other words, if the user orders to perform the setting operation, the voice command identifier 100 performs the setting operation shown in Fig. 5, and otherwise it performs the normal operation shown in Fig. 6.

[0036] Fig. 5 shows a flowchart of a “setting operation” shown in Fig. 4 according to an embodiment of the present invention. As described above, when the user ordered to perform the setting operation and the setting operation starts, each and every variable stored in the first to fourth sub-memories 300, 302, 304 and 306 is reset to have a predetermined value, for example zero (0), (step S502). Then, a total repetition count P of the setting operation, which shows how many times the setting operation will be performed for current trial, is set according to a user’s preference or a predetermined default value. And, a current repetition count q of the setting operation, which shows how many times the setting operation has been performed for current trial, is initialized to a predetermined value, for example zero ($q=0$), (step S504). The total repetition count P of the step S504 may be set to a predetermined value during its manufacturing, or may be set by the user every time the setting operation is performed.

[0037] Next, a variable k is initialized (for example, $k=0$) (step S506). The variable k shows the order of a sampled value during a predetermined setting period Δt for digitizing an analog signal. The variable k has a value in the range of zero (0) to a predetermined maximum value N, which is dependent on the storage capacity of the memory device used, the processing performance of the microprocessor 114, required accuracy of voice command identification, etc.

[0038] Then, the microprocessor 114 controls the output selecting switch 124 to couple output of the speaker 102 to the second digital-to-analog converter 122, so that a sound signal data corresponding to a pulse $\delta(t)$ having amplitude of one (1) is generated during the setting period Δt , and a sound according to the sound signal data is output from the speaker 102 (step S508).

[0039] Figs. 7a and 7b show waveforms of a pulse output during the step S508 and an electrical signal $S_{mic}(t)$ generated by the microphone 104 receiving the pulse signal, respectively. As shown in the drawing, $M(k)$ is defined to be a value of a digital signal, to which the pulse $\delta(t)$ is digitized, and then each $M(k)$ has a value of one (1)

during the setting period Δt . It is only because of the calculation simplicity to generate the pulse $\delta(t)$ as described above to have the amplitude of one (1), therefore it is also possible to generate the pulse $\delta(t)$ to have a value other than one (1) according to another embodiment. This embodiment is described later. Further, the setting period Δt is a very short period of time (i.e. several milliseconds) in practice, so there is no possibility for an audience to hear the sound resulted from the pulse $\delta(t)$.

[0040] Next, the second digital-to-analog converter 116 converts the object signal $S_{\text{command}}(t)$ into digital signals, and stores the digital signals to the fourth sub-memory 306 (step S510). At this moment, while performing the current step, the first digital-to-analog converter 116 does not generate any signal. Therefore, the object signal $S_{\text{command}}(t)$ is identical to the electrical signal $S_{\text{mic}}(t)$ from the microphone. Further, the value of the variable $D(k)$ is repeatedly acquired by performing the setting process P times, and the P values of the $D(k)$'s may be averaged. The subscript q shows the order of the acquired value of $D(k)$. This is also true to other variables. Thus, in case the setting operation is performed only once, the subscript q has no meaning. Further, the operation of converting an analog signal into digital signals is represented as a function, $Z[]$, in the drawing.

[0041] Next, a value of $D(k)$ acquired during current setting operation is accumulated to that (or those) acquired during prior setting operation(s). Next, it is determined whether or not the variable k is equal to the maximum value N , and, if the result is negative, the above described steps S510 to S514 are repeated until k becomes equal to N .

[0042] Next, it is determined whether or not the subscript q is equal to the total repetition count P (step S516), and, if the result is negative, the subscript q is increased by a predetermined unit (step S518) and the above steps S506 to S516 are repeated.

[0043] After completing the above described steps, final values of variables $D(k)$'s are divided by the total repetition count P , and then the divided values are stored in the first sub-memory 306 as environmental coefficients $C(k)$'s, respectively. The environmental coefficient $C(k)$ is based on the following Equation 2;

[Equation 2]

$$0 = D(k) - C(k) * Z[\delta(t)]$$

[0044] Here, since $Z[\delta(t)]$ is a pulse of a value known to the microprocessor 114, it may be considered to have a value of one (1) by the second digital-to-analog converter 122. Thus, it is possible to say $D(k) = C(k)$. Further, as described above, each value of $D(k)$ acquired during each setting operation is accumulated to $D(k)$ itself, and the final $D(k)$ should be divided by the total repetition count P to get an averaged value of the $D(k)$.

[0045] In case the pulse generated in the step S508 has a value A other than one (1), a value of $P \cdot A$, P multiplied by A , is calculated. Then, the final value of each $D(k)$ is divided by the value of $P \cdot A$ and the divided value of each $D(k)$ is stored in the first sub-memory 306 as the environment coefficient $C(k)$.

[0046] As described later, the $C(k)$ is multiplied by the data $M(k)$ digitized from a sound signal during a normal operation to become a sound source data for generating approximation signal $\text{Sum}(\text{Dis})$, which is an approximation of a noise signal $S_{\text{dis}}(t)$ of the Equation 1.

[0047] Steps of the setting operation are performed as described above. According to another embodiment of the present invention, steps S522 to S530 may additionally be performed to acquire more precise calculations. This is described in detail, hereinafter.

[0048] After acquiring the environment coefficient $C(k)$, the microprocessor 114 stores random data to the third sub-memory 304 as a temporary value of the variable $M(k)$, which is then used to generate sound output through speaker 102 (step S522). Next, a “normal operation”, as described in detail later, is performed (step S524) to determine whether or not the object signal $S_{\text{command}}(t)$ is substantially zero (0) (step S526). If the result of the determination of the step S526 is affirmative, the current environmental coefficient $C(k)$ is stored (step S530) and the control is returned. If negative, the current environmental coefficient $C(k)$ is corrected (step S528), and the steps S524 and S526 are repeated.

[0049] As described above, since the environmental coefficient $C(k)$ may be corrected during the normal operation, the environmental coefficient $C(k)$ having an initial value due to the initial environment may have new value due to changed environment. For example, if the system is a television, existence of an audience may require new value of the environmental coefficient $C(k)$. Or, change of the number of audience(s) may be regarded as change of the environment, which make the reflection

characteristics different. So, it may be required for the environmental coefficient $C(k)$ to be corrected to have a new value corresponding to the new environment in this case, also.

[0050] It is preferable to store the environmental coefficient $C(k)$ in a non-volatile memory, as described above. It is not required to re-acquire the environmental coefficient $C(k)$ when the system power is off and on again with the non-volatile memory storing the environmental coefficient $C(k)$ if the environment has not been changed. However, as described above, if the amount of power consumption is not important, a volatile memory may be used, but in this case the setting operation is performed after the system power is on again.

[0051] Fig. 6 shows a flowchart of the “normal operation” shown in Fig. 4 according to an embodiment of the present invention. As described above with reference to Fig. 4, it is preferable to automatically perform the normal operation (step S406) if the setting operation (step S404) is not performed.

[0052] Now, referring Fig. 6 again, after the operation starts, the microprocessor 114 loads the environmental coefficient $C(k)$ to the fast second sub-memory 302 from the slow first sub-memory 300, and the loaded environmental coefficient $C(k)$ in the second sub-memory 302 is designated as “ $C^{RAM}(k)$ ” (step S602). At this moment, the clocking variable T may be initialized (i.e. $T=0$), which is described later.

[0053] Next, the microprocessor 114 receives volume data C' from the audio signal generator 108, multiplies the environmental coefficient $C^{RAM}(k)$ loaded to the second sub-memory 302 by the volume data C' to acquire weighted environmental coefficient $C'(k)$ (step S604).

[0054] Next, the sound signal $S_{org}(t)$ from the audio signal generator 108 is converted into digital data M during a predetermined sampling period (step S606). The converted digital data M is stored in the third sub-memory 304 as data $M(k)$ by Que operation (step S608). The steps S606 and S608 are repeated during the sampling period, and every converted digital data at each sampling time point t_k is stored in the third sub-memory 304 as the data $M(k)$.

[0055] Next, a pseudo-distortion signal $Sum(Dis)$ is calculated using the $M(k)$ in the third sub-memory 304 and the weighted environment coefficient $C'(k)$ according to the following Equation 3 (step S610).

[Equation 3]

$$Sum(Dis) = \sum_{k=0}^N C'(k)M(k)$$

[0056] Here, N is an upper limit, which is based on an assumption that the sampling period and the sampling frequency are equal to those used for the setting operation.

[0057] Now, with reference to Fig. 8, the physical meaning of the pseudo-distortion signal Sum(Dis) is described in detail. Fig. 8 shows waveforms of the sound signal $S_{org}(t)$ output from the audio signal generator 108 during the normal operation and the electrical signal $S_{mic}(t)$ received and generated from the microphone 104. If the sampling period is from t_0 to t_6 and the present time point is t_7 , various sound signals, which are output from the speaker 102 from t_0 to t_7 and distorted by various environmental variables via various paths (i.e. paths d_1 to d_6 as shown in Fig. 1), are superposed and input to the microphone 104. Thus, the electrical signal $S_{mic}(t_7)$ generated by the microphone 104 at the present time point t_7 includes superposed signals of the user's command signal and the distorted signals. Since the superposed signals of the distorted signals reflect cumulative effects of the environmental variables, the pseudo-distorted signals $Sum(Dis)_{t=7}$ at the present time point t_7 may be represented as the following Equation 4;

[Equation 4]

$$\begin{aligned} Sum(Dis)_{t=7} &= \sum_{k=0}^6 C'(k)M(k) \\ &= [C'(0)M(0)+C'(1)M(1)+C'(2)M(2)+C'(3)M(3) \\ &\quad +C'(4)M(4)+C'(5)M(5)+C'(6)M(6)] \end{aligned}$$

[0058] Next, the first digital-to-analog converter 116 converts the pseudo-distortion signal Sum(Dis) into an analog signal (step S612), and the adder 118 subtracts the converted pseudo-distortion signal from the electrical signal $S_{mic}(t)$ to generate the object signal $S_{command}(t)$ which is to be recognized by the voice recognizer 110 (step S614).

[0059] By performing the above described steps, the possibility for the voice recognizer 110 to perform false recognition is substantially decreased to zero (0) even though the sound output from the speaker 102 includes sounds similar to voice

commands, which may be recognized by the voice recognizer 110, because the pseudo-distortion signal Sum(Dis) corresponding to the sounds similar to voice commands is subtracted from the signals input to the microphone 104.

[0060] The normal operation of the voice command identifier 100 according to an embodiment of the present invention is completed by completing the above steps. However, even during the above described normal operation, the environment may be change from one during the setting operation by a user's movement or entrance of a new audience. Therefore, it may be preferable to perform the above described steps S 502 to S520 of the setting operation shown in Fig. 5 during the normal operation at an every predetermined time. In this case, steps S616 to S628 as shown in Fig. 6 may be additionally performed, as described hereinafter.

[0061] It is determined whether or not the clocking variable T initialized in the step S602 becomes to be equal to a predetermined clocking value (i.e. 10) (step S616). The clocking variable T is used to indicate elapsed time for performing the normal operation of steps S602 to S614, and may easily be embodied by system clock in practice. Further, the predetermined clocking value is set to perform the setting operation at an every predetermined time, for example 10 seconds, and may be set by a manufacturer or a user.

[0062] If the determination result of the step S616 shows that the current value of the clocking variable T is not yet equal to the predetermined clocking value, the value of the clocking variable is increased by a unit value (i.e. one(1)) as a unit time (i.e. one (1) second) has elapsed (step S618), and the normal operation of the steps S604 to S616.

[0063] However, if the determination result of the step S616 shows that the current value of the clocking value T is equal to the predetermined clocking value, the microprocessor 114 controls the output selecting switch 124 to select the second digital-to-analog converter 122 and to couple it to the speaker 102, and to initialize the value of the clocking variable T (i.e. $T=0$), again.

[0064] Next, the microprocessor 144 controls the speaker 102 not to generate any sound (step S622). This is to wait until remaining noise around the system disappears.

[0065] Next, after a predetermined time period for waiting for the noise to disappear, the microprocessor 144 detects the electrical signal $S_{mic}(t)$ from the microphone 104 for another predetermined time period (step S624), and determines whether or not any noise is included in the detected electrical signal $S_{mic}(t)$ (step S626). By doing this, it is possible to determine whether or not external noise is input into the

microphone 104 because it is difficult to acquire normal environmental coefficient $C(k)$ under the presence of the external noise. In case the determination result of the step S626 shows that external noise is detected, the present setting operation may be canceled to return control to the step S604, and the normal operation is continued.

[0066] However, if the external noise is not detected, the setting operation of steps S502 to S520 is performed (step S628).

Figs. 9a and 9b respectively show waveforms of an output signal output from the speaker 102 when the renewal setting operation (steps S616 to S628) during the normal operation is performed and one output when it is not performed. As shown in the drawings, it is preferable that the step S622 is started during the first Δt period and maintained for the second Δt period, the steps S624 and S626 are performed during the second Δt period, and the step S628 is performed during the third Δt period. Of course, actual duration of the Δt period may be adjusted according to the embodiments.

[0067] Fig. 9c shows a waveform of an output signal output from the speaker 102 while the waveform shown in Fig. 9a is output two (2) times. As shown in the drawing, actual duration of the time period, or $3\Delta t$, for performing the renewal setting operation is very short (i.e. several milliseconds), so the user can not notice the performance of the renewal setting operation.

[0068] According to one embodiment of the present invention, it is possible to identify a user's voice command from sound signals reflected and re-input and to allow a credible voice recognition in a system having its own sound source. Further, it is also possible to achieve a real time voice recognition due to substantial reduction of amount of calculation.

[0069] While the above description has pointed out novel features of the invention as applied to various embodiments, the skilled person will understand that various omissions, substitutions, and changes in the form and details of the device or process illustrated may be made without departing from the scope of the invention. Therefore, the scope of the invention is defined by the appended claims rather than by the foregoing description. All variations coming within the meaning and range of equivalency of the claims are embraced within their scope.